Karlsruher Institut für Technologie

Teilchenphysik 2 — W/Z/Higgs an Collidern

Sommersemester 2019

Matthias Schröder und Roger Wolf | Vorlesung 8

INSTITUT FÜR EXPERIMENTELLE TEILCHENPHYSIK (ETP)

Termine: Bisher

Date	Room	Type	Topic
Wed Apr 24.	Kl. HS B	LE 01	1. Organisation and introduction: particle physics at colliders + W/Z/H history
Tue Apr 30.	30.23 11/12	—	*no class*
Wed May 01.	Kl. HS B	—	*no class*
Tue May 07.	30.23 11/12	LE 02	2.1 Gauge theory & 2.2 The electroweak sector of the SM I
Wed May 08.	Kl. HS B	LE 03, EX 01	2.3 Discovery of the W and Z bosons & EX gauge theories
Tue May 14.	30.23 11/12	LE 04	2.4 The Higgs mechanism
Wed May 15.	Kl. HS B	EX 02	Exercise “SM Higgs mechanism”
Tue May 21.	30.23 11/12	—	*no class*
Wed May 22.	Kl. HS B	LE 05	2.5 The electroweak sector of the SM II (Higgs mechanism + Yukawa couplings)
Tue May 28.	30.23 11/12	SP 01	Specialisation of 2.4 and 2.5
Wed May 29.	Kl. HS B	LE 06	3.1 From theory to observables & 3.2 Reconstruction + analysis of exp. data
Tue Jun 04.	30.23 11/12	EX 03	Exercise “Trigger efficiency measurement”
Wed Jun 05.	Kl. HS B	LE 07	3.3 Measurements in particle physics (part 1)
Tue Jun 11.	30.23 11/12	EX 04	Exercise on statistical methods
Wed Jun 12.	Kl. HS B	LE 08	3.3 Measurements in particle physics (part 2)
Tue Jun 18.	30.23 11/12	SP 02	Specialisation “Limit setting”
Wed Jun 19.	Kl. HS B	SP 03	Specialisation “Unfolding”
Tue Jun 25.	30.23 11/12	LE 09	4.1 Determination of SM parameters
Wed Jun 26.	Kl. HS B	LE 10	4.2 Measurement and role of W/Z bosons at the LHC
Tue Jul 02.	30.23 11/12	EX 05	Paper seminar “Z pole measurements”
Wed Jul 03.	Kl. HS B	LE 11	4.3 Processes with several W/Z bosons
Tue Jul 09.	30.23 11/12	EX 06	Paper seminar Higgs
Wed Jul 10.	Kl. HS B	LE 12	5.1 Discovery and first measurements of the Higgs boson
Tue Jul 16.	30.23 11/12	EX 07	Exercise “Machine learning in physics analysis”
Wed Jul 17.	Kl. HS B	LE 13	5.2 Measurement of couplings and kinematic properties
Tue Jul 23.	30.23 11/12	EX 08	Presentations: results of ML challenge
Wed Jul 24.	Kl. HS B	LE 14	5.3 Search for Higgs physics beyond the SM & 5.4 Future Higgs physics

Termine: NEU

Date	Room	Type	Topic
Wed Apr 24.	Kl. HS B	LE 01	1. Organisation and introduction: particle physics at colliders + W/Z/H history
Tue Apr 30.	30.23 11/12	—	*no class*
Wed May 01.	Kl. HS B	—	*no class*
Tue May 07.	30.23 11/12	LE 02	2.1 Gauge theory & 2.2 The electroweak sector of the SM I
Wed May 08.	Kl. HS B	LE 03, EX 01	2.3 Discovery of the W and Z bosons & EX gauge theories
Tue May 14.	30.23 11/12	LE 04	2.4 The Higgs mechanism
Wed May 15.	Kl. HS B	EX 02	Exercise “SM Higgs mechanism”
Tue May 21.	30.23 11/12	—	*no class*
Wed May 22.	Kl. HS B	LE 05	2.5 The electroweak sector of the SM II (Higgs mechanism + Yukawa couplings)
Tue May 28.	30.23 11/12	SP 01	Specialisation of 2.4 and 2.5
Wed May 29.	Kl. HS B	LE 06	3.1 From theory to observables & 3.2 Reconstruction + analysis of exp. data
Tue Jun 04.	30.23 11/12	EX 03	Exercise “Trigger efficiency measurement”
Wed Jun 05.	Kl. HS B	LE 07	3.3 Measurements in particle physics (part 1)
Tue Jun 11.	30.23 11/12	EX 04	Exercise on statistical methods
Wed Jun 12.	Kl. HS B	LE 08	3.3 Measurements in particle physics (part 2)
Tue Jun 18.	30.23 11/12	SP 02	Specialisation “Limit setting”
Wed Jun 19.	Kl. HS B	LE 09	4.1 Determination of SM parameters
Tue Jun 25.	30.23 11/12	SP 03	Specialisation “Unfolding”
Wed Jun 26.	Kl. HS B	LE 10	4.2 Measurement and role of W/Z bosons at the LHC
Tue Jul 02.	30.23 11/12	EX 05	Paper seminar “Z pole measurements”
Wed Jul 03.	Kl. HS B	LE 11	4.3 Processes with several W/Z bosons
Tue Jul 09.	30.23 11/12	EX 06	Paper seminar Higgs
Wed Jul 10.	Kl. HS B	LE 12	5.1 Discovery and first measurements of the Higgs boson
Tue Jul 16.	30.23 11/12	EX 07	Exercise “Machine learning in physics analysis”
Wed Jul 17.	Kl. HS B	LE 13	5.2 Measurement of couplings and kinematic properties
Tue Jul 23.	30.23 11/12	EX 08	Presentations: results of ML challenge
Wed Jul 24.	Kl. HS B	LE 14	5.3 Search for Higgs physics beyond the SM & 5.4 Future Higgs physics

Analysis Chain

Nature $\overset{?}{\longleftrightarrow}$ Theory

$\Downarrow$ (Nature) $\Downarrow$ (Theory)

Detector: data recording
calibrated digitised data
online selection (trigger)

MC simulation
physics process
detector signals

$\Downarrow$

Physics object **reconstruction**
Event **selection**

$\Downarrow$

Statistical analysis: **results**
Comparison with theory

3. From Theory to Experiment (and Back)

3.1 From theory to observables
- Cross-section calculation: basic picture
- Fermion propagator and perturbation theory
- Scattering matrix and Feynman rules

3.2 Reconstruction of experimental data
- Reminder: accelerators and particle detectors
- Trigger
- Reconstruction of physics objects

3.3 Measurements in particle physics
- Parameter estimation
- Hypothesis testing
- Search for new physics (exclusion limits)

3.4 Monte Carlo simulation

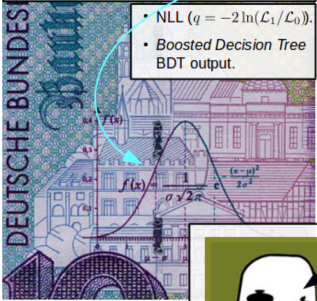
3.3 Measurements in particle physics

Statistics and Stochastic (Probability)

Test statistic:

$\Omega^n \to \mathbb{R}: \qquad x \to f(x)$

- NLL ($q = -2\ln(\mathcal{L}_1/\mathcal{L}_0)$).
- *Boosted Decision Tree* BDT output.



Probability (density) function:

$\Omega^n \to [0,1] \subset \mathbb{R}: \qquad x \to \mathcal{P}(x)$

- Problem of statistics is usually *ill-defined*.
- Deduce *truth* from shadows in Platon's cave...

From R. Wolf

Statistics and Particle Physics

Karlsruher Institut für Technologie

Experiment

- All measurements are derived from **counting experiments**
- **Millions of billions** of particle collision events each year
- Each event **independent** from and (to many extents) identical to all others

Theory

- Nature intrinsically stochastic
- QM wave functions interpreted as **probability density functions**
- Event-by-event simulation using **Monte Carlo** methods

Particle physics experiments perfect application of statistics

Literature

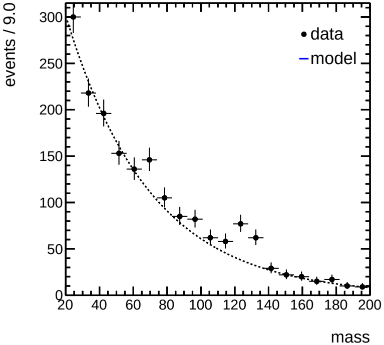
- R. J. Barlow, “Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences”, The Manchester Physics Series **highly recommended**
- G. Cowan, “Statistical Data Analysis”, Oxford Science Publications
- Review “Statistics”, Chin. Phys. C, 40, 100001 (2016) **very good review**
- The ATLAS and CMS Collaborations, “Procedure for the LHC Higgs boson search combination in Summer 2011”, CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11 **expert document**

Likelihood

- **Likelihood** $\boxed{\mathcal{L}(n_{\text{obs}}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_{\text{obs}} = \{n_{\text{obs},i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$

  → **function of model parameters and observed data**
- Example
  - Data: observed events in 20 bins
  - Model: known SM process (background)

$$\mathcal{L}(n_{\text{obs}}; b) = \prod_i \text{Poisson}(n_{\text{obs},i}, b)$$



Likelihood

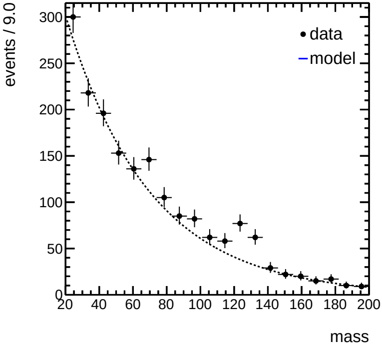
- **Likelihood** $\boxed{\mathcal{L}(n_{\text{obs}}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_{\text{obs}} = \{n_{\text{obs},\,i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$

  → **function of model parameters and observed data**

- Example
  - Data: observed events in 20 bins
  - Model: known SM process (background)

$$\mathcal{L}(n_{\text{obs}}; b) = \prod_i \frac{(n_{\text{pred},i})^{n_{\text{obs},i}}}{n_{\text{obs},i}!} \mathrm{e}^{-n_{\text{pred},i}}$$

$$n_{\text{pred},i}(b) = \underbrace{b \cdot \mathrm{e}^{-\alpha x_i}}_{\text{background}}$$

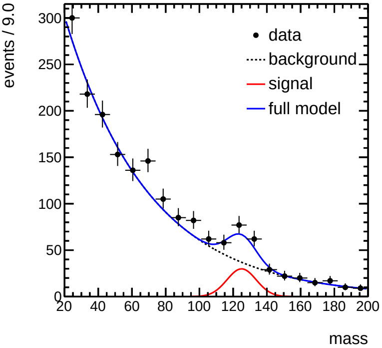


Likelihood

- **Likelihood** $\boxed{\mathcal{L}(n_{\text{obs}}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_{\text{obs}} = \{n_{\text{obs},\,i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$
  - → **function of model parameters and observed data**
- Example
  - Data: observed events in 20 bins
  - Model: known SM process (background) + signal

$$\mathcal{L}(n_{\text{obs}}; b, s) = \prod_i \frac{(n_{\text{pred},i})^{n_{\text{obs},i}}}{n_{\text{obs},i}!} e^{-n_{\text{pred},i}}$$

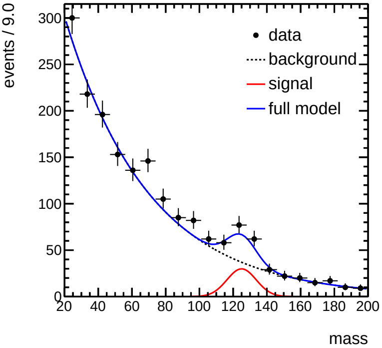
$$n_{\text{pred},i}(b, s) = \underbrace{b \cdot e^{-\alpha x_i}}_{\text{background}} + \underbrace{s \cdot e^{-(m-x_i)^2}}_{\text{signal}}$$



Likelihood

- **Likelihood** $\boxed{\mathcal{L}(n_\text{obs}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_\text{obs} = \{n_{\text{obs},\,i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$

  $\rightarrow$ **function of model parameters and observed data**
- $\mathcal{L}$ is **not the probability** of a model!
  - Also not the probability of observing the data given the model
- $\mathcal{L}$ is the product of pdfs of the model, used to quantify the compatibility of the data with the model



3.3.1. Parameter estimation

Parameter Estimation

- Which parameters of a given model describe best the data?
- **Maximum likelihood (ML) method** ("maximum likelihood fit"): Which parameter values lead to the maximum of $\mathcal{L}$?

→ **Maximum likelihood estimators $\hat{\mathbf{k}}$** of the true values

$$\boxed{\max\left[\mathcal{L}(n_{\text{obs}}, k)\right] = \mathcal{L}(n_{\text{obs}}, \hat{k})}$$

- Technically: minimisation of negative log-likelihood

$$\text{NLL}(n_{\text{obs}}, k) = -2 \ln \mathcal{L}$$

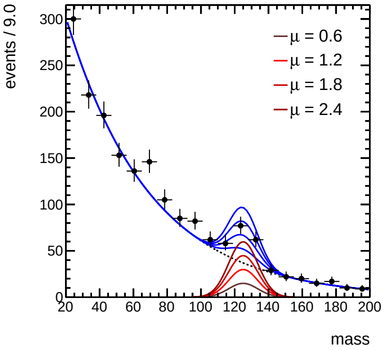
  - Logarithm is monotonic: retains minimum
  - Turns product of pdfs into sum: easier to minimise
- Arbitrarily complex problem, vast amount of techniques and literature
  - e. g. ATLAS+CMS Higgs couplings combination: $\approx$ 4250 parameters
  - e. g. CMS tracker-alignment problem: $\mathcal{O}(10^6)$ parameters

Maximum-Likelihood Method

- Example: which *signal-strength modifier* μ describes best the data?
  - Common in Higgs physics: $\mu = (\sigma \cdot \mathcal{B})/(\sigma_{\text{SM}} \cdot \mathcal{B}_{\text{SM}})$

$$n_{\text{pred},i}(\mu) = \underbrace{b \cdot e^{-\alpha x_i}}_{\text{background}} + \underbrace{\mu \cdot s_{\text{SM}} \cdot e^{-(m-x_i)^2}}_{\text{signal}}$$

$$\mathcal{L}(n_{\text{obs}};\mu) = \prod_i \frac{(n_{\text{pred},i})^{n_{\text{obs},i}}}{n_{\text{obs},i}!} e^{-n_{\text{pred},i}}$$



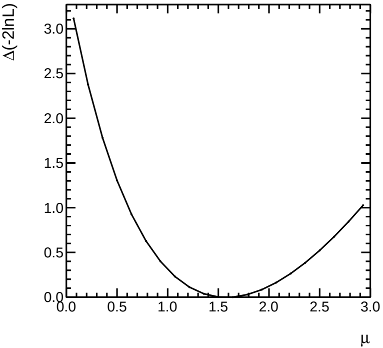
→ $\hat{\mu}$: maximises $\mathcal{L}(n_{\text{obs}};\mu)$

Maximum-Likelihood Method

- Example: which *signal-strength modifier* μ describes best the data?
  - Common in Higgs physics: $\mu = (\sigma \cdot \mathcal{B})/(\sigma_{\text{SM}} \cdot \mathcal{B}_{\text{SM}})$

$$n_{\text{pred},i}(\mu) = \underbrace{b \cdot e^{-\alpha x_i}}_{\text{background}} + \underbrace{\mu \cdot s_{\text{SM}} \cdot e^{-(m-x_i)^2}}_{\text{signal}}$$

$$\mathcal{L}(n_{\text{obs}};\mu) = \prod_i \frac{(n_{\text{pred},i})^{n_{\text{obs},i}}}{n_{\text{obs},i}!} e^{-n_{\text{pred},i}}$$



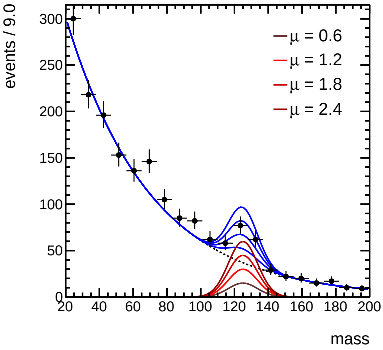
→ $\hat{\mu}$: minimises $-2 \ln \mathcal{L}(n_{\text{obs}};\mu)$

Maximum-Likelihood Method

- Example: which *signal-strength modifier* μ describes best the data?
  - Common in Higgs physics: $\mu = (\sigma \cdot \mathcal{B})/(\sigma_{\text{SM}} \cdot \mathcal{B}_{\text{SM}})$
- Important case: **Gaussian distributed measurements**
  - e. g. approximation of Poisson for large number of events (in practice > 10)
- NLL becomes

$$-2 \ln \mathcal{L} = \sum_i \frac{(n_{\text{obs},i} - n_{\text{pred},i})^2}{n_{\text{pred},i}} + \text{const}$$

$$\chi^2 \equiv \sum_i \frac{(n_{\text{obs},i} - n_{\text{pred},i})^2}{n_{\text{pred},i}}$$

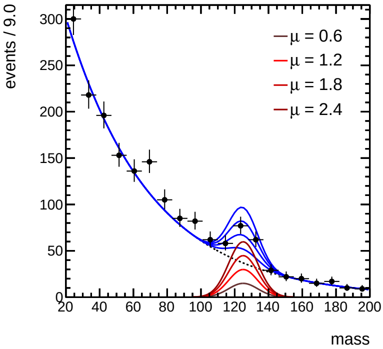


Maximum-Likelihood Method

- Example: which *signal-strength modifier* μ describes best the data?
  - Common in Higgs physics: $\mu = (\sigma \cdot \mathcal{B})/(\sigma_{\text{SM}} \cdot \mathcal{B}_{\text{SM}})$

ML method obtains parameter values most compatible with the data *for a given model* — **a ML fit does not find the correct model**!

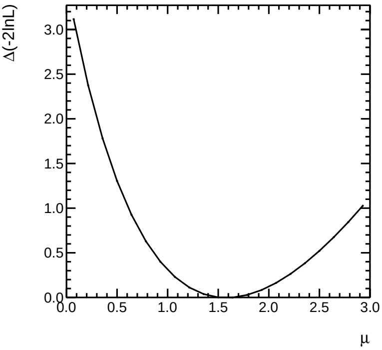
ML estimator is function of the observed data



Maximum-Likelihood Method

- Uncertainty on $\hat{\mu}$ from scan of NLL = $-2\ln\mathcal{L}$ around minimum
- **Uncertainty due to fluctuations of data**: "statistical uncertainty"
- For Gaussian pdfs: parabola, standard deviation σ follows from

$$\Delta(\text{NLL}) = \text{NLL}(\hat{\mu} \pm r \cdot \sigma) - \text{NLL}(\hat{\mu}) = r$$



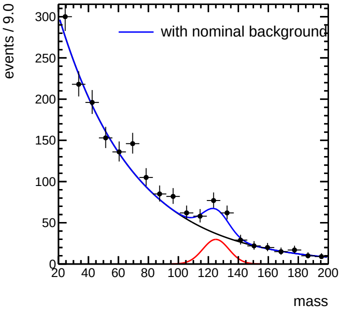
- Other cases can often be approximated by Gaussian case (and true for $n \to \infty$)
- For strongly asymmetric NLL
  - Often asymmetric intervals quoted
  - Better: variable transformation such that NLL symmetric

Uncertainty vs. Error

- **Uncertainty** reflects degree of precision with which one can deduce a parameter value from the data if one does everything correctly
  - Statistical uncertainties stem from the inherent stochastic nature of the data and finite number of observed events (can not be eliminated, only minimised)
  - Systematic uncertainties stem e. g. from the limited knowledge of the precision of the detector or approximations in theory calculations
  - Uncertainties can (in principle) be quantified
- In contrast, **errors are mistakes**
  - e. g. a lose cable or a wrong method
  - Errors can (and should) be eliminated
- In context of statistical data analysis, we mean uncertainties
  - NB: “Error bars” denote uncertainties!

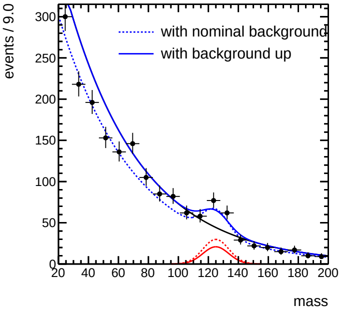
Incorporation of Systematic Uncertainties

- Often, background (and signal) models subject to systematic uncertainties
  - Some (theory or experimental) parameters not exactly known, e. g. cross section or trigger efficiency
- For example, background normalisation *b* not precisely known but with some uncertainty



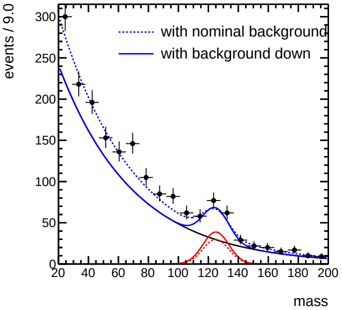
Incorporation of Systematic Uncertainties

- Often, background (and signal) models subject to systematic uncertainties
  - Some (theory or experimental) parameters not exactly known, e. g. cross section or trigger efficiency
- For example, background normalisation b not precisely known but with some uncertainty $\rightarrow$ affects determination of signal
  - Generally leads to **larger uncertainty on $\hat{\mu}$**



Incorporation of Systematic Uncertainties

- Often, background (and signal) models subject to systematic uncertainties
  - Some (theory or experimental) parameters not exactly known, e. g. cross section or trigger efficiency
- For example, background normalisation *b* not precisely known but with some uncertainty $\rightarrow$ affects determination of signal
  - Generally leads to **larger uncertainty on $\hat{\mu}$**

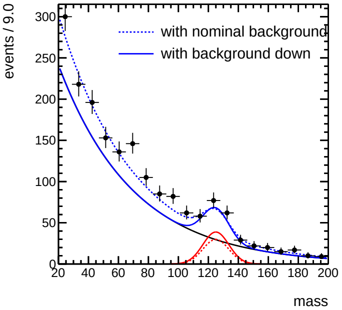


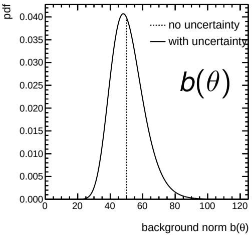
Incorporation of Systematic Uncertainties

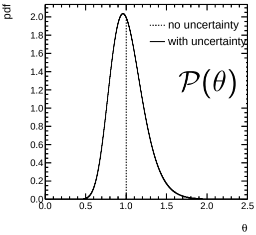
- Incorporated into likelihood via **nuisance parameters** θ

$$\mathcal{L}(\mathbf{n}_{\text{obs}};\mu,\theta) = \prod_i \mathcal{P}(n_{\text{obs}\,i};\mu,\theta)\cdot\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)$$

- Background normalisation becomes function of θ: $b \rightarrow b(\theta)$
  - θ: assumed true value, parameter of the fit
  - $\tilde{\theta}$: best knowledge, e. g. estimate from independent measurement, distributed as $\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)$





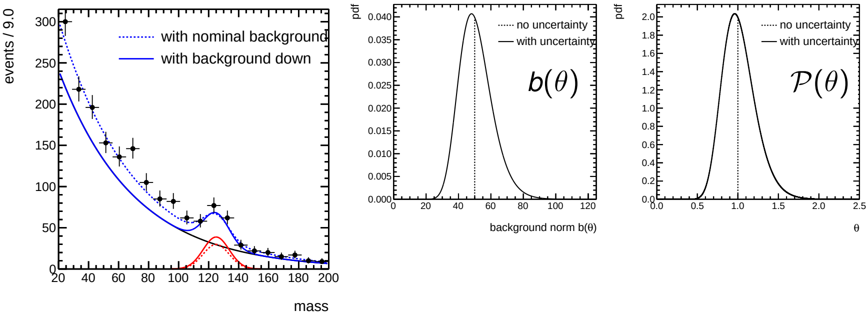


Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\mathcal{L}(\mathbf{n}_{\text{obs}}; \mu, \theta) = \prod_i \mathcal{P}(n_{\text{obs}\,i}; \mu, \theta) \cdot \mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)$$

- Background normalisation becomes function of θ: $b \to b(\theta)$
- ML fit can adjust θ to a value different than $\tilde{\theta}$ to achieve better description of data but at the cost of reducing value of $\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)$

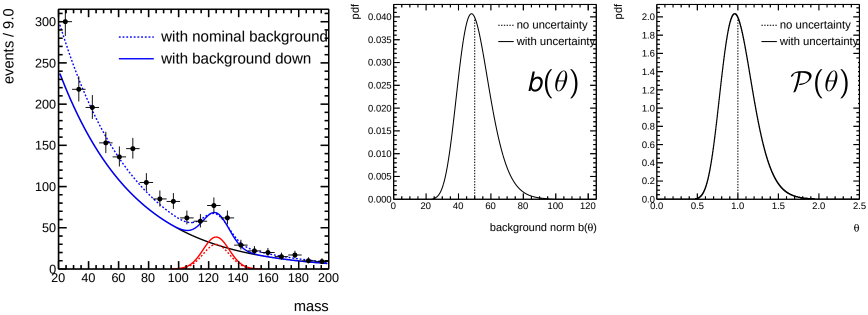


Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\mathcal{L}(\mathbf{n}_{\text{obs}};\mu,\theta)=\prod_i \mathcal{P}(n_{\text{obs}\,i};\mu,\theta)\cdot\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)$$

- Nomenclature
  - μ: *parameter of interest* (POI)
  - θ: *nuisance parameter* (NP)
    generally, (anti-)correlated with μ: increased uncertainty on $\hat{\mu}$

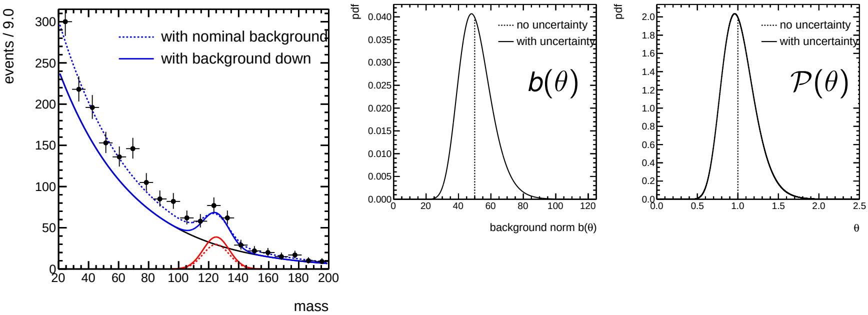


Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\mathcal{L}(\mathbf{n}_{\mathrm{obs}};\mu,\theta) = \prod_i \mathcal{P}(n_{\mathrm{obs}\,i};\mu,\theta)\cdot\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)$$

- Nomenclature
  - μ: *parameter of interest* (POI)
  - θ: *nuisance parameter* (NP)
    generally, (anti-)correlated with μ: smaller sensitivity



Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\mathcal{L}(\mathbf{n}_{\text{obs}}; \mu, \theta) = \prod_i \mathcal{P}(n_{\text{obs}\,i}; \mu, \theta) \cdot \mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)$$

- Nomenclature
  - μ: *parameter of interest* (POI)
  - θ: *nuisance parameter* (NP)
- $\mathcal{L}$ function of μ and θ, but only interested in μ
  $\rightarrow$ can rewrite θ as function of μ

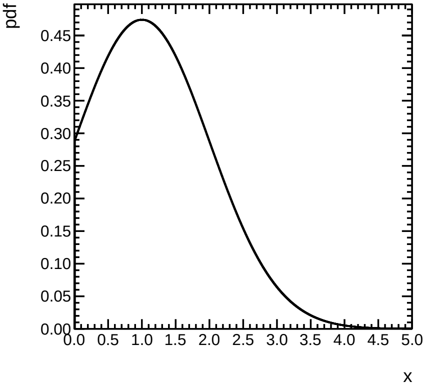
Profile likelihood $\mathcal{L}_{\text{p}}(\mu) \equiv \mathcal{L}(\mu, \hat{\theta}(\mu))$

$\hat{\theta}(\mu)$ maximises $\mathcal{L}$ for given μ (“profiled values” of θ)

- $\mathcal{L}_{\text{p}}$ in practice computationally advantageous
  - e. g. need only to consider μ instead of full parameter space (μ, θ) when computing uncertainties

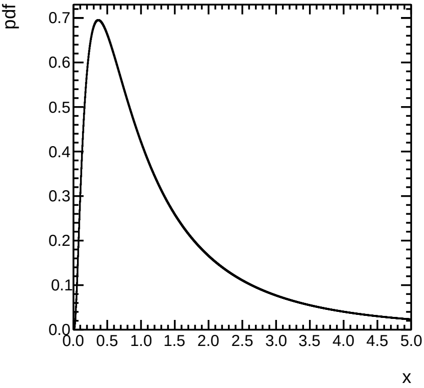
Typical Choices for $\mathcal{P}_\theta(\theta|\tilde{\theta})$

(truncated) Gaussian



$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

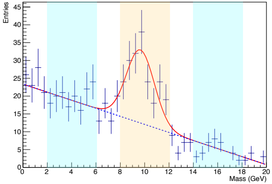
Log-normal



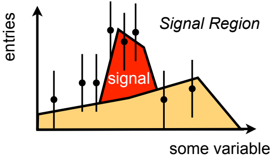
$$\frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{1}{2}(\frac{\ln(x)-\mu}{\sigma})^2}$$

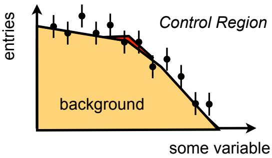
Sidebands and Control Regions

sidebands



signal/control region





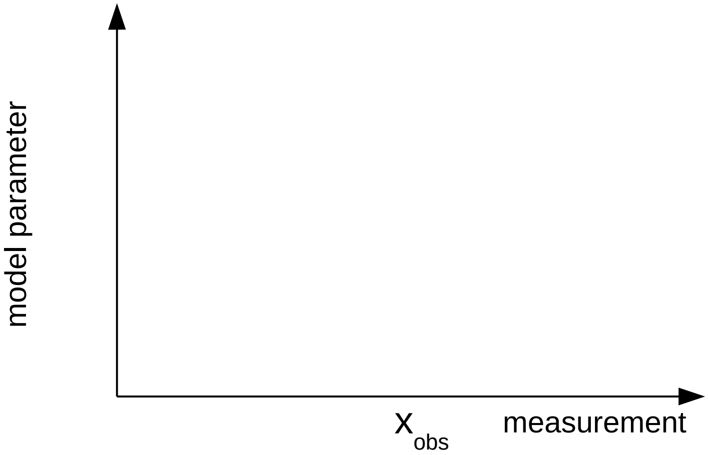
- Signal-depleted sideband/control region
  - Determine rate of background process
  - Determine shape of background processes
  - Constrain uncertainties on background prediction
- Can be incorporated into likelihood fit via **nuisance parameters**: **simultaneous determination of POI and background parameters**

Confidence Intervals

- Experiment to determine value of some parameter x measures x_{obs}
  - e. g. from maximum likelihood estimate, tag-and-probe measurement
- Want to quote "uncertainty": interval that reflects statistical precision of measurement
  - Can be estimate of standard deviation from likelihood fit
  - More generally, in particular if non-Gaussian $\mathcal{P}$: **confidence interval**
- **Confidence interval** covers on average the true value with a given probability
  - e. g. *90 % confidence level*: if experiment repeated many time, the interval covers the true value in 90 % of the cases
  - Careful: this is not the probability of the true value to lie within the interval (there is only one true value either within or not)!

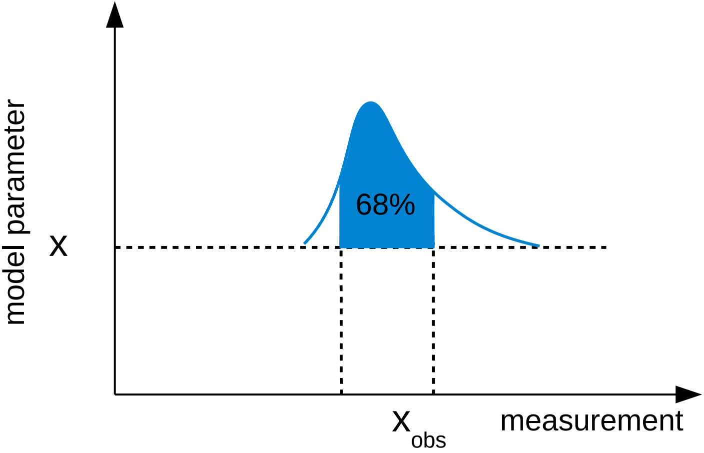
Confidence Intervals

Neyman construction

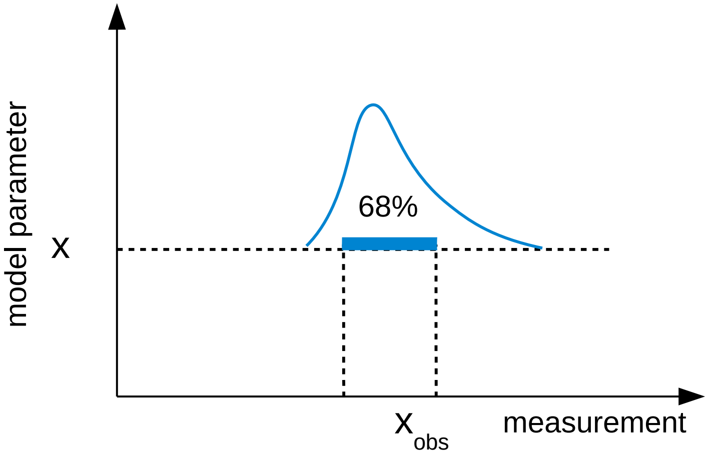


Confidence Intervals

Neyman construction

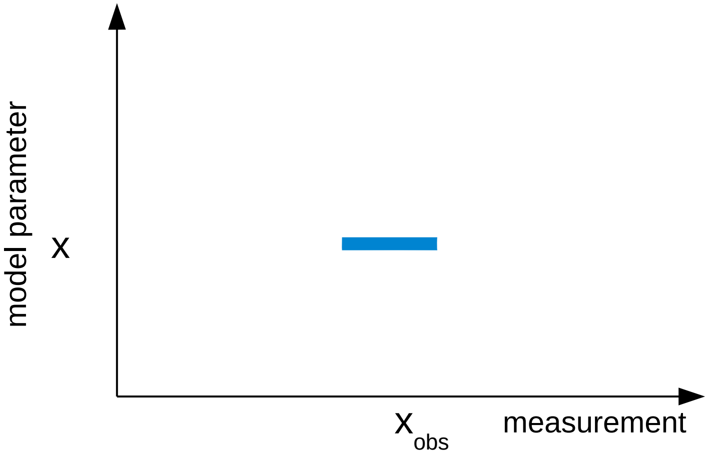


Confidence Intervals
Neyman construction



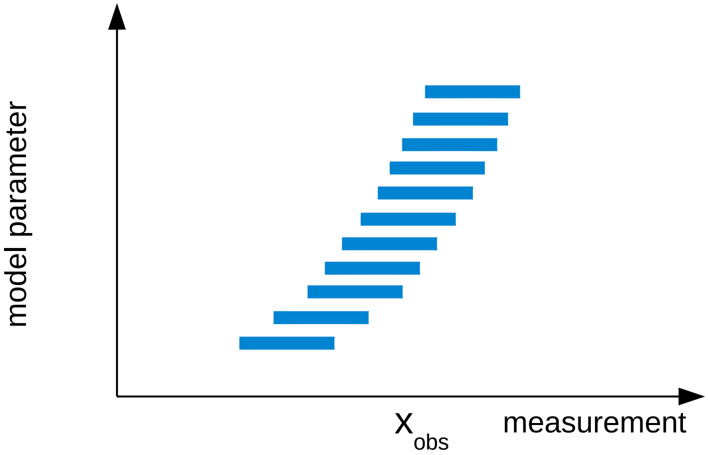
Confidence Intervals

Neyman construction



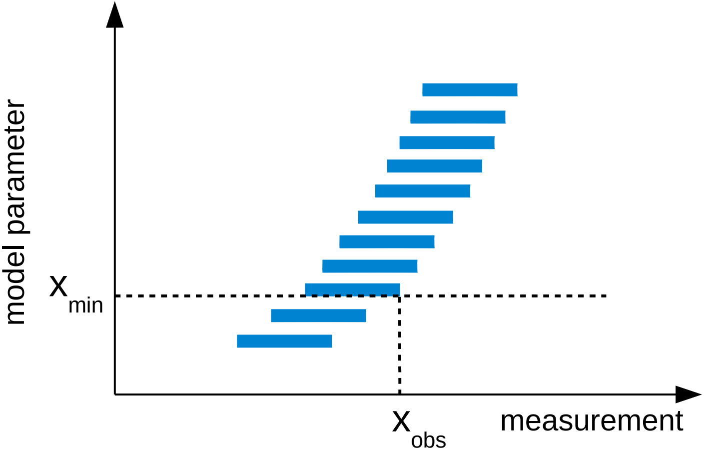
Confidence Intervals

Neyman construction



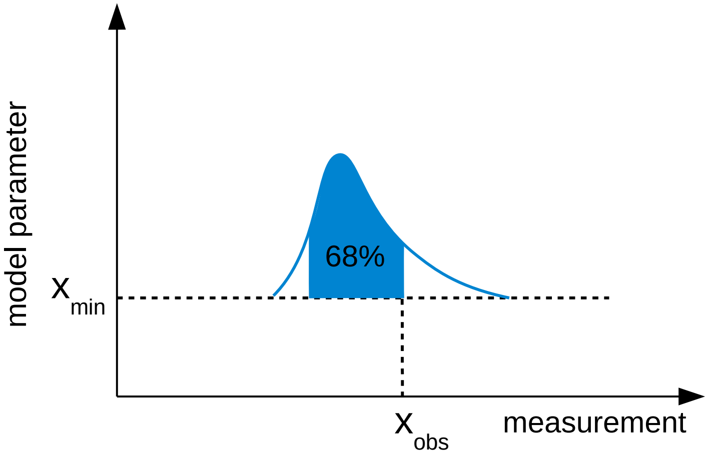
Confidence Intervals

Neyman construction



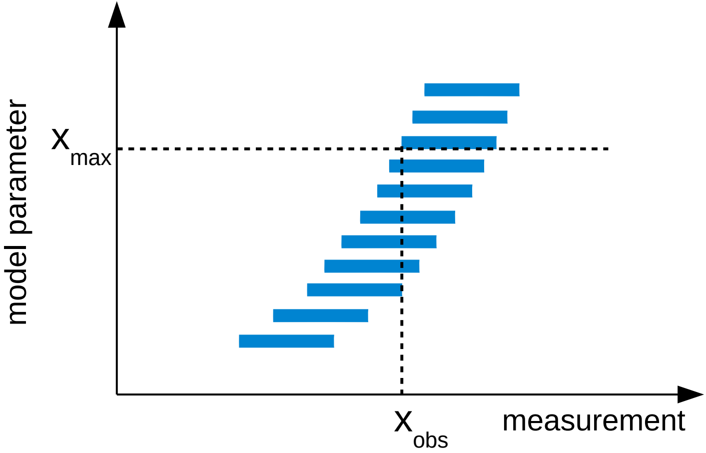
Confidence Intervals

Neyman construction



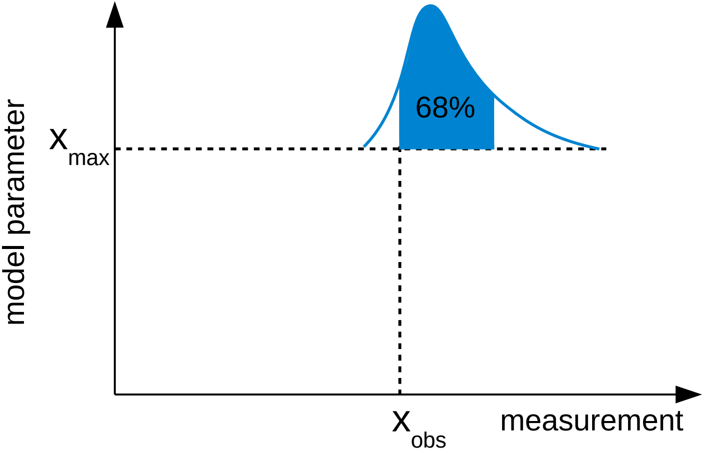
Confidence Intervals

Neyman construction



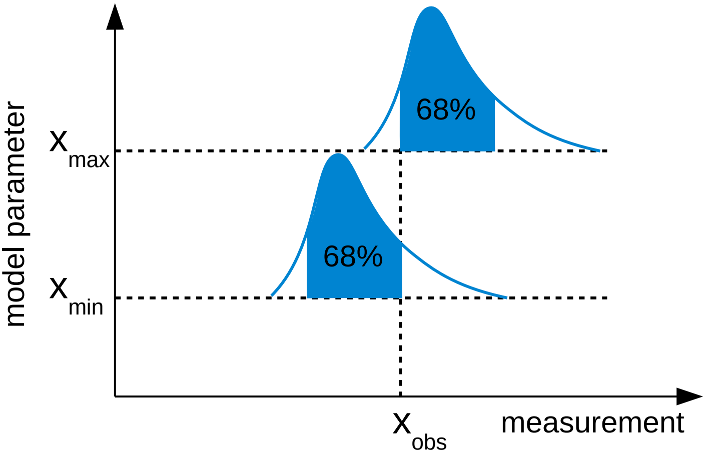
Confidence Intervals

Neyman construction



Confidence Intervals

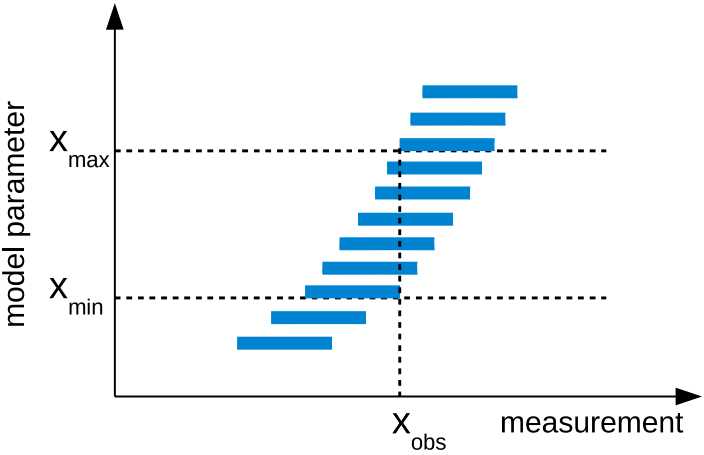
Neyman construction



Confidence interval is $[x_{\mathrm{min}}, x_{\mathrm{max}}]$

Confidence Intervals

Neyman construction



Confidence interval is $[x_{min}, x_{max}]$

Application: Uncertainty of Efficiency

(see Exercises No. 3)

- Efficiency: **numberator = subset of denominator**
  $\rightarrow$ fully correlated, don't use Gaussian error propagation!
- Good description: **binomial uncertainties**
  - Given true efficiency ϵ
  - Draw from a population of N events: expect $\langle n \rangle = \epsilon N$ events on average
  - Variance of n

$$V[n] \equiv \sigma_n^2 = N\epsilon(1-\epsilon)$$

  - But don't know true efficiency: replace with estimator $\hat{\epsilon}$ from measurement (random sample)

$$\epsilon \rightarrow \hat{\epsilon} = \frac{n}{N}, \qquad \hat{\sigma}_n^2 = N\hat{\epsilon}(1-\hat{\epsilon})$$

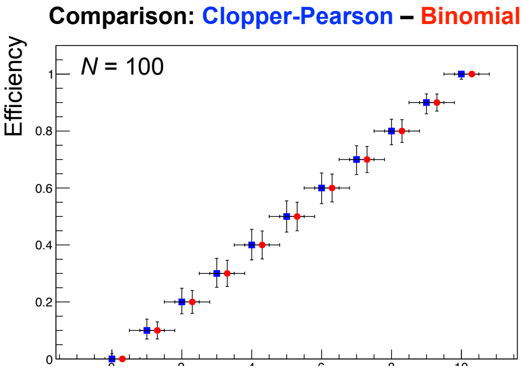
  - Uncertainty on efficiency estimator

$$\boxed{\sigma_{\hat{\epsilon}}^2 = \frac{1}{N^2}\hat{\sigma}_n^2 = \frac{\hat{\epsilon}(1-\hat{\epsilon})}{N}}$$

Application: Uncertainty of Efficiency

(see Exercises No. 3)

- Problem with binomial uncertainty from *measured* efficiency: variance $\rightarrow 0$ for $\hat{\epsilon} \rightarrow 0, 1$
- Solution: construction of proper **68 % confidence level intervals**

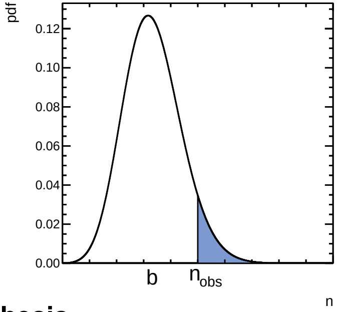


3.3.2. Hypothesis testing

p Value

- How well is the data described by my hypothesis (model)?
- For example, **how likely that observed peak in data just upward fluctuation** of the background, i. e. no Higgs boson present?
- Assume simple counting experiment
  - *b*: expected number of background events (=model)
  - n_{obs}: number of observed events
- "*p* value": **probability of upward fluctuation as large as or larger than observed in data**

$$p \equiv \mathrm{P}(n \geq n_{\text{obs}} | b) = \int_{n_{\text{obs}}}^{\infty} \mathrm{d}n \, \mathcal{P}(n|b)$$



p value is **not the probability of a hypothesis**

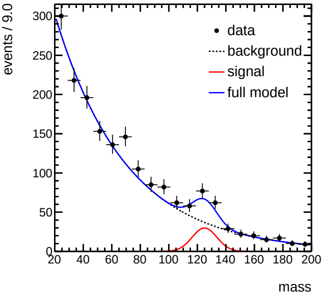
p quantifies level of (dis-)agreement between model and data:

$\rightarrow$ judgement call whether to keep model or reject it

Test Statistic

- Typically problems not as simple as a single measured quantity
  - Multiple channels/measurements, e. g. binned distribution
  - Prediction depends on several parameters, e. g. POI, nuisance params.
- Formally: use appropriate **test statistic** $t : \mathbb{R}^D \to \mathbb{R}$
  - In general any function that **combines relevant information** from an experiment, e. g. number of observed events per bin, **into one single number** reflecting the agreement between data and hypothesis
  - Likelihood is example for test statistic
- t can be used to compute p value for complex models

$$p = \int_{t_{\text{obs}}}^{\infty} \mathcal{P}(t)\,\mathrm{d}t$$

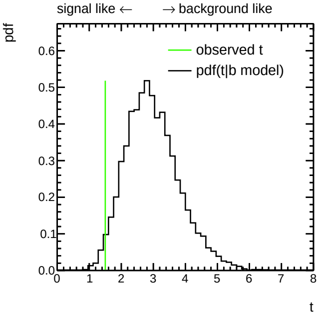


Test Statistic

- Typically problems not as simple as a single measured quantity
  - Multiple channels/measurements, e. g. binned distribution
  - Prediction depends on several parameters, e. g. POI, nuisance params.
- Formally: use appropriate **test statistic** $t : \mathbb{R}^D \to \mathbb{R}$
  - In general any function that **combines relevant information** from an experiment, e. g. number of observed events per bin, **into one single number** reflecting the agreement between data and hypothesis
  - Likelihood is example for test statistic
- t can be used to compute p value for complex models

$$p = \int_{t_{\text{obs}}}^{\infty} \mathcal{P}(t)\,\mathrm{d}t$$

  - Requires pdf $\mathcal{P}(t)$ of the test statistic t: typically from simulation (*toy data*)

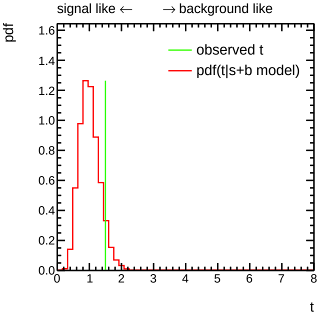


Test Statistic

- Typically problems not as simple as a single measured quantity
  - Multiple channels/measurements, e. g. binned distribution
  - Prediction depends on several parameters, e. g. POI, nuisance params.
- Formally: use appropriate **test statistic** $t : \mathbb{R}^D \to \mathbb{R}$
  - In general any function that **combines relevant information** from an experiment, e. g. number of observed events per bin, **into one single number** reflecting the agreement between data and hypothesis
  - Likelihood is example for test statistic
- t can be used to compute p value for complex models

$$p = \int_{t_{\text{obs}}}^{\infty} \mathcal{P}(t)\,\mathrm{d}t$$

  - Requires pdf $\mathcal{P}(t)$ of the test statistic t: typically from simulation (*toy data*)

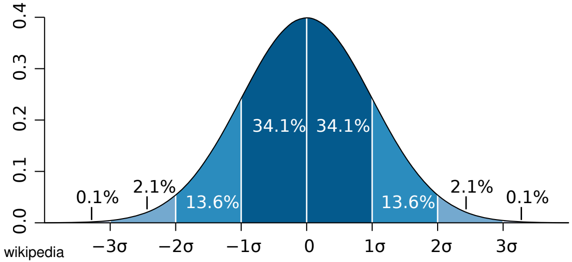


Significance

- Often, p value converted into equivalent **significance Z**:
  upward fluctuation from 0 by Z of normal-distributed variable corresponding to same p value
  - Corresponds to Z standard deviations σ of the Gaussian distribution

$$Z = \Phi^{-1}(1-p)$$

Φ: cumulative (=quantile) function of normal distribution



- **Convention** to classify effects by significance
  - 3σ: *evidence* for signal (0.3% chance of background fluctuation)
  - 5σ: *discovery* of signal (0.00006% chance of background fluctuation)

Distinguishing Hypotheses

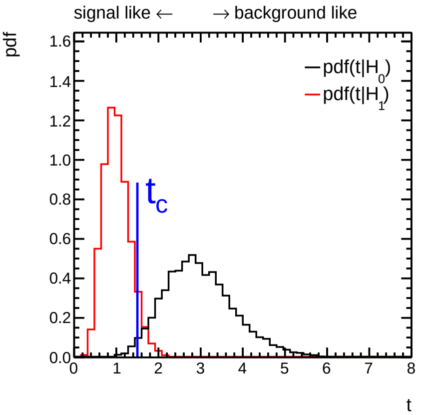
- **Cannot determine whether a hypothesis is true** (frequentist)
- But can define rules how to reject hypothesis in favour of an alternative hypothesis
- Can determine probability of wrong choice: how often wrong choice is made would the experiment be repeated very often

Distinguishing Hypotheses

- Classification problem: how to interpret outcome of an experiment?
- Want to distinguish between two alternative hypotheses, e. g.
  - H_0: there are only background events, e. g. no Higgs boson
  - H_1: there is a signal, e. g. a Higgs boson, contribution to the data

- Define **test statistic *t***, e. g. $\mathcal{L}$
- Construct pdf $\mathcal{P}(t|H_i)$ of t under H_0 and H_1
  - In reality often from simulation
- How compatible is t_{obs} with H_i?

→ Set a critical value t_c and reject H_0 in favour of H_1 if $t_{obs} < t_c$



Distinguishing Hypotheses

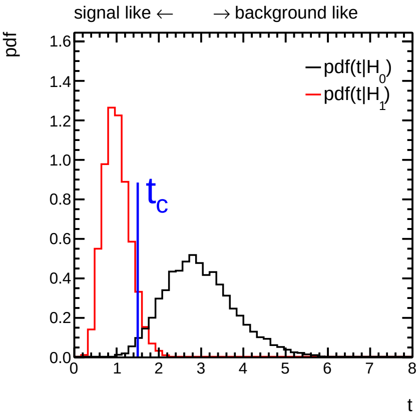
- Classification problem: how to interpret outcome of an experiment?
- Want to distinguish between two alternative hypotheses, e. g.
  - H_0: there are only background events, e. g. no Higgs boson
  - H_1: there is a signal, e. g. a Higgs boson, contribution to the data

- Types of errors
  - I: reject H_0 although it is true
  - II: accept H_0 although H_1 is true

$$\text{I: } P(t < t_c | H_0) = \int_{-\infty}^{t_c} \mathrm{d}t \, \mathcal{P}(t|H_0) \equiv \alpha$$
$$\text{II: } P(t > t_c | H_1) = \int_{t_c}^{\infty} \mathrm{d}t \, \mathcal{P}(t|H_1) \equiv \beta$$

α : significance of test

$1 - \beta$: power of test



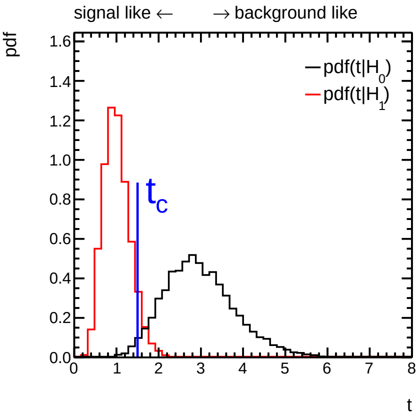
Distinguishing Hypotheses

- Classification problem: how to interpret outcome of an experiment?
- Want to distinguish between two alternative hypotheses, e. g.
  - H_0: there are only background events, e. g. no Higgs boson
  - H_1: there is a signal, e. g. a Higgs boson, contribution to the data

- Types of errors
  - I: reject H_0 although it is true
  - II: accept H_0 although H_1 is true

$$\text{I: } \mathrm{P}(t < t_c | H_0) = \int_{-\infty}^{t_c} \mathrm{d}t\, \mathcal{P}(t|H_0) \equiv \alpha$$
$$\text{II: } \mathrm{P}(t > t_c | H_1) = \int_{t_c}^{\infty} \mathrm{d}t\, \mathcal{P}(t|H_1) \equiv \beta$$

By choice of t_c (choice of α): association to one or the other hypothesis performed "**at the confidence level α**"



Difference Between α and *p* Value?

- In one sense, no difference
  - Both are $\int_x^\infty \mathrm{d}x' \, \mathcal{P}(x')$
- But **conceptually very different**
  - α is computed before one sees the data: predefined property of the test
  - *p* depends on the data (property of the data) and is a random variable

Neyman-Pearson Lemma

- When performing a test between two hypotheses (models) H_0 and H_1, the **likelihood ratio test**, which rejects H_0 in favour of H_1 if

$$\boxed{Q = \frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}} > Q_c} \quad \text{with } \mathrm{P}(Q > Q_c | H_0) = \alpha$$

  is the **most powerful test** at a significance level α
- The test statistic Q is called **likelihood ratio**
- For a number of reasons, usually

$$\boxed{q = -2 \ln Q}$$

Test Statistic at the LHC

- Used test statistic in particle-physics experiments evolved with time
- Hypothesis test example
  - H_0: background-only hypothesis, e. g. SM without Higgs boson
  - H_1: Higgs boson with fixed mass and signal strength μ present in data
- At the LHC commonly: profile likelihood ratio test statistics

$$q_\mu = -2 \ln \frac{\mathcal{L}(n_{\mathrm{obs}}; \mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(n_{\mathrm{obs}}; \hat{\mu}, \hat{\theta})}$$

  - Nuisance parameters are profiled in nominator
  - Global maximum of $\mathcal{L}$ under (μ, θ) as denominator with $0 \leq \hat{\mu} \leq \mu$
  - Allows usage of certain approximations when computing $\mathcal{P}(q_\mu | H_i)$ ("asymptotic formulae" based on Wilks and Wald theorem[1])

[1] G. Cowan "Asymptotic formulae for likelihood-based tests of new physics", Eur. Phys. J. C71:1554 (2011)

3.3.3. Search for new physics (exclusion limits)

Excluding Parameters

- Assume measurement with a given sensitivity: no signal observed
- How much signal can “hide” in the bkg. fluctuations (+uncertainty)?
- **How large could a signal be at most?**

Excluding Parameters

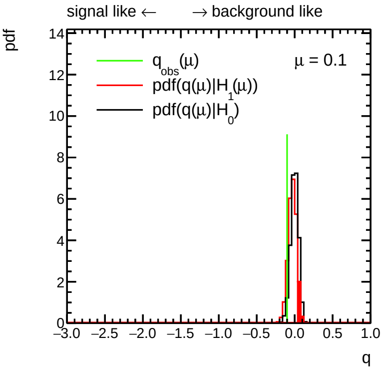
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2\ln\frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$ for which H_1 would be rejected at significance level α?

$$\alpha = \int_{q_{\mathrm{obs}}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL_{s+b}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\mathrm{obs}} = q_c$



Excluding Parameters

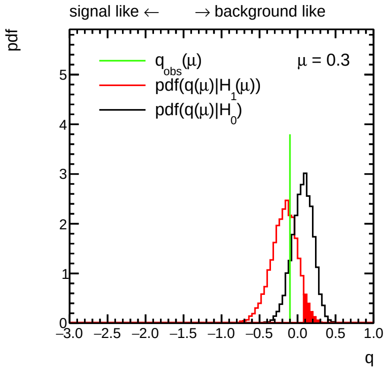
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2\ln\frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$ for which H_1 would be rejected at significance level α?

$$\alpha = \int_{q_{\rm obs}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL_{s+b}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\rm obs} = q_c$



Excluding Parameters

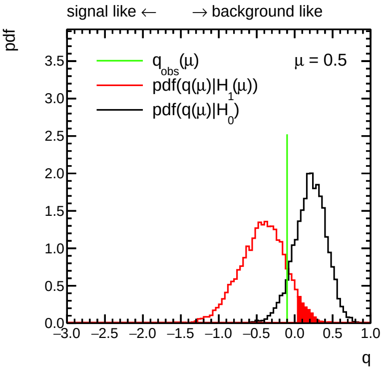
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2 \ln \frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$ for which H_1 would be rejected at significance level α?

$$\alpha = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q \, \mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{\mathrm{s+b}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\text{obs}} = q_c$



Excluding Parameters

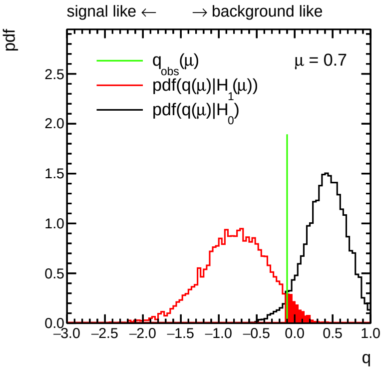
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2 \ln \frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$ for which H_1 would be rejected at significance level α?

$$\alpha = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q \, \mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{\mathrm{s+b}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\text{obs}} = q_c$



Excluding Parameters

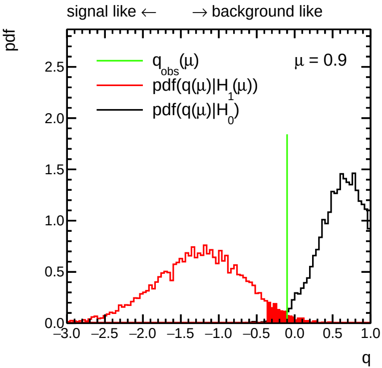
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2\ln\frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$ for which H_1 would be rejected at significance level α?

$$\alpha = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q\,\mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \text{CL}_{\text{s+b}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\text{obs}} = q_c$



Excluding Parameters

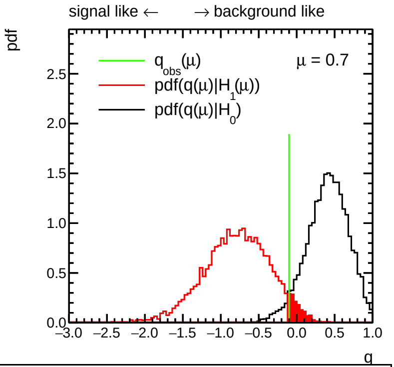
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2 \ln \frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$ for which H_1 would be rejected at significance level α?

$$\alpha = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q \, \mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{\mathrm{s+b}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\text{obs}} = q_c$



$\mu_{1-\alpha}$ is called **upper limit** on μ at confidence level C.L. $= 1 - \alpha$

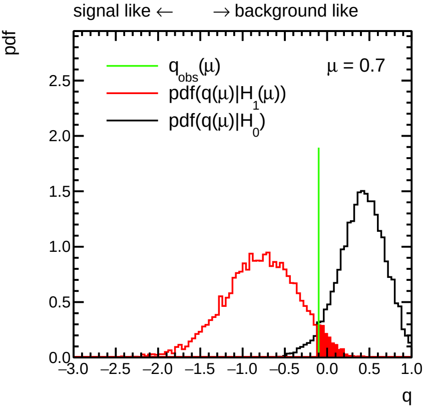
In particle physics usually 95 % C.L. limit

(Observed) Upper Limit: Interpretation

- "Maximal signal that we would still reject"
- 95 % C.L. upper limit on μ: largest value of μ that would still be rejected in a test with significance 5% given the data
  - NB: limit is a **function of the data** (depends on q_{obs})!
- μ for which $\text{CL}_{\text{s+b}} = 0.05$:

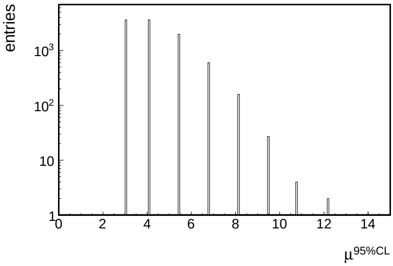
$$0.05 = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q \, \mathcal{P}(q(\mu_{95})|H_1)$$

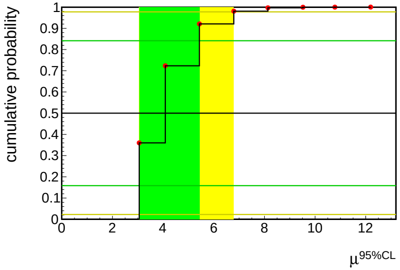
- Upper **limit covers true value ($\mu_{\text{true}} < \mu_{95}$) with probability C.L. $=$ 95 %**
  - If the experiment is repeated many times, μ_{95} would be larger than μ_{true} in 95 % of the cases
- Still **5 % chance of wrong exclusion**, i. e. that $\mu_{\text{true}} > \mu_{95}$



Expected Limit

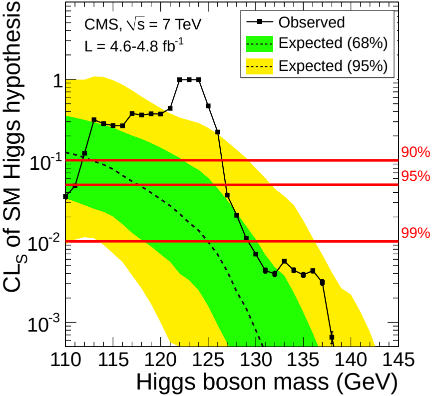
- **Estimate what observed limit would look like in case of no signal**
- Obtained e. g. from *toy dataset*
  - Sample toy data for q under background-only hypothesis from $\mathcal{P}(q|H_0)$
  - Treat each as observation and compute μ_{95} limit
  - Obtain quantiles from distribution of all μ_{95}
- **Expected limit = median of μ_{95} distribution**
  - 16 and 84% quantiles: 68% confidence interval
  - 2.5 and 97.5% quantiles: 95% confidence interval
    $\rightarrow$ “Brazilian band” plots





Before the Higgs-Boson Discovery

- Combination of Higgs-boson search results by CMS [Phys.Lett. B710 (2012) 26]

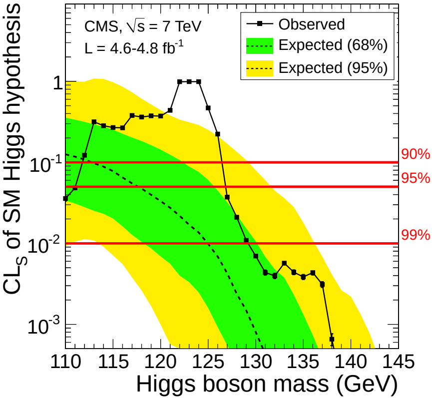


- Tested hypotheses
  - H_0: no Higgs boson
  - H_1: SM Higgs boson
- Test statistic q evaluated for SM Higgs boson of different mass ($\mu = 1$ in each case)

- Excluding a SM Higgs boson at 95% CL with masses
  - $m_H > 118$ GeV expected (from toy data under H_0)
  - $m_H > 127$ GeV observed (from real data)

Before the Higgs-Boson Discovery

- Combination of Higgs-boson search results by CMS [Phys.Lett. B710 (2012) 26]



- Tested hypotheses
  - H_0: no Higgs boson
  - H_1: SM Higgs boson
- Test statistic q evaluated for SM Higgs boson of different mass ($\mu = 1$ in each case)

- Observed exclusion weaker than expected (smaller mass range)
- Around $m_H = 125\,\text{GeV}$, q_{obs} differs significantly from expectation: **indication that H_0 is wrong!**